Contents lists available at ScienceDirect

Indian Heart Journal

journal homepage: www.elsevier.com/locate/ihj

Original Article

A novel approach for the prediction of treadmill test in cardiology using data mining algorithms implemented as a mobile application

A. Jerline Amutha[a,b,*], R. Padmajavalli[c,e], D. Prabhakar[d]

[a] Research Scholar, Department of Computer Science, Bharathiar University, Coimbatore, India
[b] Assistant Professor, Department of Computer Science, Women's Christian College, Chennai, India
[c] Head, Department of Computer Applications, Bhaktavatsalam Memorial College for Women, Korattur, Chennai, India
[d] Consultant Cardiologist, Ashwin Clinic, AG 25, Annanagar, Chennai, 600040, India
[e] Research Supervisor, Department of Computer Science, Bharathiar University, Coimbatore, Chennai, India

ARTICLE INFO

ABSTRACT

Objective: To develop a mobile app called "TMT Predict" to predict the results of Treadmill Test (TMT), using data mining techniques applied to a clinical dataset using minimal clinical attributes. To prospectively test the results of the app in realtime to TMT and correlate with coronary angiogram results.
Methods: In this study, instead of statistics, data mining approach has been utilized for the prediction of the results of TMT by analyzing the clinical records of 1000 cardiac patients. This research employed the Decision Tree algorithm, a new modified version of K-Nearest Neighbor (KNN) algorithm, K-Sorting and Searching (KSS). Furthermore, curve fitting mathematical technique was used to improve the Accuracy. The system used six clinical attributes such as age, gender, body mass index (BMI), dyslipidemia, diabetes mellitus and systemic hypertension. An Android app called "TMT Predict" was developed, wherein all three inputs were combined and analyzed. The final result is based on the dominating values of the three results. The app was further tested prospectively in 300 patients to predict the results of TMT and correlate with Coronary angiography.
Results: The accuracy of predicting the result of a TMT using data mining algorithms, Decision Tree and K-Sorting & Searching (KSS) were 73% and 78%, respectively. The mathematical method curve fitting predicted with 82% accuracy. The accuracy of the mobile app "TMT Predict", improved to 84%. Age-wise analysis of the results show that the accuracy of the app dips when the age is more than 60 years indicating that there may be other factors like retirement stress that may have to be included. This gives scope for future research also. In the prospective study, the positive and negative predictive values of the app for the results of TMT and coronary angiogram were found to be 40% and 83% for TMT and 52% and 80% for coronary angiogram. The negative predictive value of the app was high, indicating that it is a good screening tool to rule out coronary artery heart disease (CAHD).
Conclusion: "TMT Predict" is a simple user-friendly android app, which uses six simple clinical attributes to predict the results of TMT. The app has a high negative predictive value indicating that it is a useful tool to rule out CAHD. The "TMT Predict" could be a future digital replacement for the manual TMT as an initial screening tool to rule out CAHD.

1. Introduction

All over the world, cardiovascular diseases account for a considerable percentage of untimely deaths. A statistical study conducted by WHO in 2012 revealed that, around 17.5 million people had died due to coronary artery heart disease (CAHD) around the world.[1] In India, nearly 45 million people suffer from cardiac diseases,[2] out of which one in four Indians die of CAHD every year.[3] Inadequate awareness and lack of seriousness often result in the loss of precious lives. Periodical screening and risk factor modification are mandatory to minimize this epidemic of CAHD.

The earlier studies[2] found that the non-modifiable risk factors such as age, gender and family history were the main reasons

* Corresponding author.
  E-mail addresses: jerlineamutha77@yahoo.com (A. Jerline Amutha), padmahari2002@yahoo.com (R. Padmajavalli), prabhud19@gmail.com (D. Prabhakar).

responsible for the heart disease. But subsequent studies[3] revealed that modifiable risk factors, namely, the lack of exercise, high fat intake, overweight, stress, smoking, alcohol, diabetes and systemic hypertension, were equally responsible for the development of heart diseases. As a result, the combination of non-modifiable and modifiable risk factors are identified as important parameters to be evaluated to predict CAHD.[4]

Exercise treadmill test (TMT) is an important screening test[5] to detect provocable ischemia. TMT is time consuming and requires physical stress. There is also a small element of risk in case of serious silent CAHD. To simplify and safely perform a screening procedure, data mining and the curve fitting mathematical procedure were applied to develop a mobile application called "TMT Predict" as a replacement for TMT. This app aimed at predicting the result of the TMT without the person actually undergoing a TMT.

The methodology for prediction includes data mining algorithms such as Decision Tree, K-Sorting & Searching (KSS) which is a new modification in K-Nearest Neighbor (KNN) and the mathematical curve fitting method.

2. Methods

At the global level, development of methods and models to improve the accuracy of diagnostic techniques is an ongoing process and a lot of research is still underway.[6] Automatic diagnosis of heart disease by taking into consideration, a number of clinical attributes (more than 13)[7,8] using data mining algorithms and evaluating its Accuracy have been done since 2001.[9]

The data mining strategies like association, clustering, classification and prediction are widely used in virtually every field.[10,11]
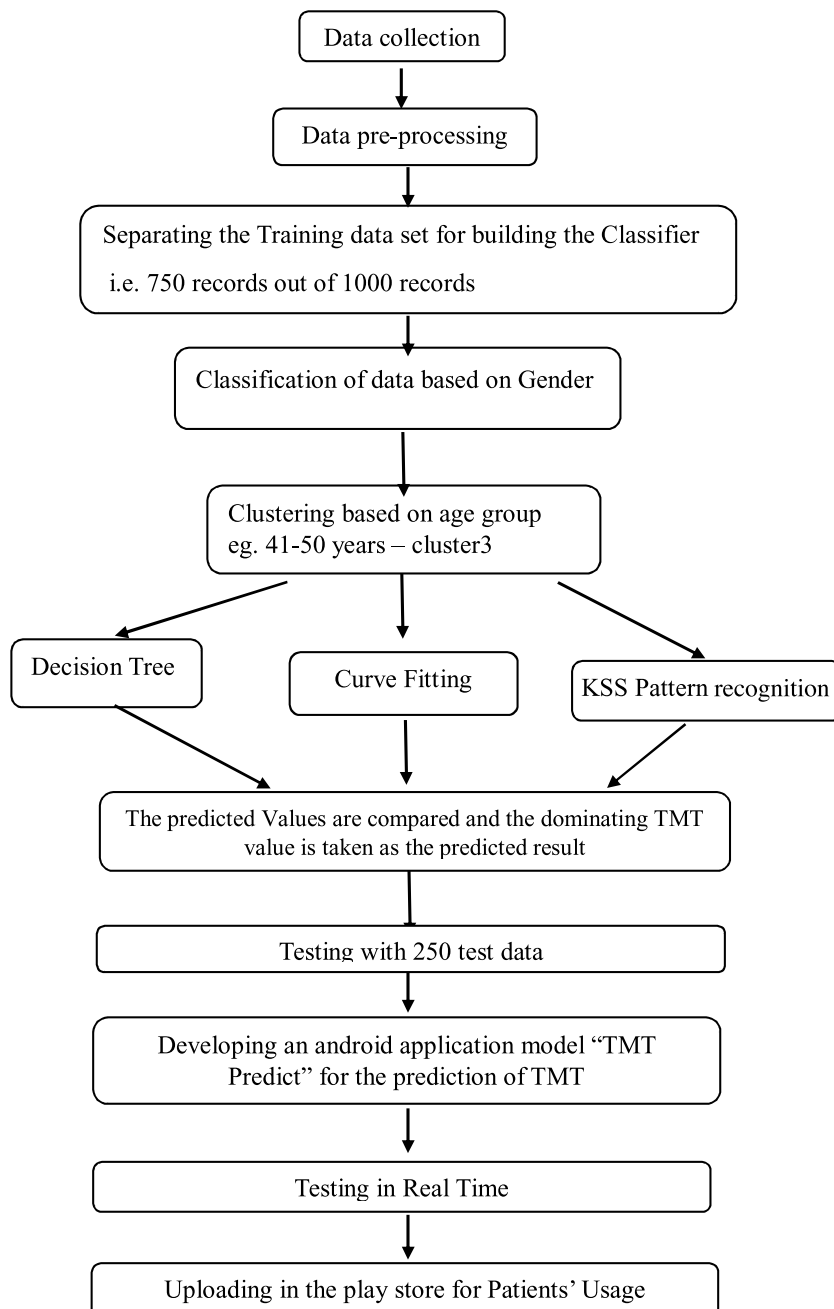

Fig. 1. Work flow of the Preparation of the Model.

Within the medical field, the classification and prediction of various diseases have frequently been analyzed. A technically flexible and strong combination of data mining and machine learning, provide a better platform for the analysis of a large volume of data.

In accordance with the same, an android mobile application called "TMT Predict" as shown in Fig. 4, was developed to predict the results of TMT. This app utilizes data mining methodologies with a minimum number of six clinical attributes. The attributes of age, sex, height and weight are entered in the app whereas BMI is calculated by the app. If the patient is known to have diabetes, hypertension or dyslipidemia, this can be entered as yes or no. There is another option to enter the values also wherein the app recognizes systemic hypertension (if BP is >139/89 mmHg), diabetes mellitus (if fasting blood sugar > 125 mg/dl or postprandial blood sugar > 199 mg/dl), and dyslipidemia (HDL < 40 mg/dl in males, HDL < 50 mg/dl in females, LDL > 70 mg/dl in diabetes, LDL > 100 mg/dl in SHT, or LDL > 130) in others.

2.1. Work flow

As illustrated in Fig. 1, the first step is data collection followed by data pre-processing, classification and clustering. Then, the data is ready for the application of algorithms to predict the results of TMT.

Two data mining algorithms, Decision Tree and KSS were applied on the processed data. In addition, a mathematical procedure called as curve fitting was also applied.

2.1.1. Data Description

The cardiology clinical dataset was acquired from a cardiology center at Chennai, where TMT was performed as a test for provocable ischemia in asymptomatic individuals without CAHD. The patient case details were electronically recorded and contain demographic details, risk factor assessment and all the test results including details of treatment and follow-up. The data is stored in cloud as MS-Excel spread sheets.

2.1.2. Data Pre-Processing

The data taken for the current study was pre-processed to eliminate inconsistent, noisy and redundant data. From this processed data set, 1000 TMT records were separated for analysis.

As discussed in paper,[12] initially all the clinical attributes were considered for classification, but, analysis of the impact of each attribute on TMT helped in identifying the most significant attributes. The six most significant attributes were age, sex, BMI, diabetes, dyslipidemia and systemic hypertension. Ultimately the most significant attributes were finalized using the entropy for each attribute as given in Eq. (1).

$$Entropy = -(p \times \log(p) + (1-p)\log(1-p)) \qquad (1)$$

where p represents positive cases and (1 – p) represents negative cases.

Further, the Boolean attributes such as gender, diabetes, hypertension and dyslipidemia were converted into numeric representation, for example, the gender attribute has a value of one for male and zero for female.

2.1.3. Data Analysis

In order to develop the "TMT Predict" mobile app, as per the data mining approach,[12] the TMT dataset was divided into a training set of 750 records, and a test set of 250 records. The training set was used to formulate a method for the prediction by considering the dependency of the attributes and the test data was used to test the formulated method.

The preliminary analysis of the most significant attributes identified "gender" as one of the prime factors in the diagnosis of the heart disease. This led to the classification of the entire training data set into male and female Classes.[12]

Subsequent analysis of data revealed that, each age group (such as age group ≤ 30, 31–40, 41–50, 51–60 and above 60) had a different behaviour on TMT result for each gender. This led to the clustering of data into different groups by means of their 'age' category.
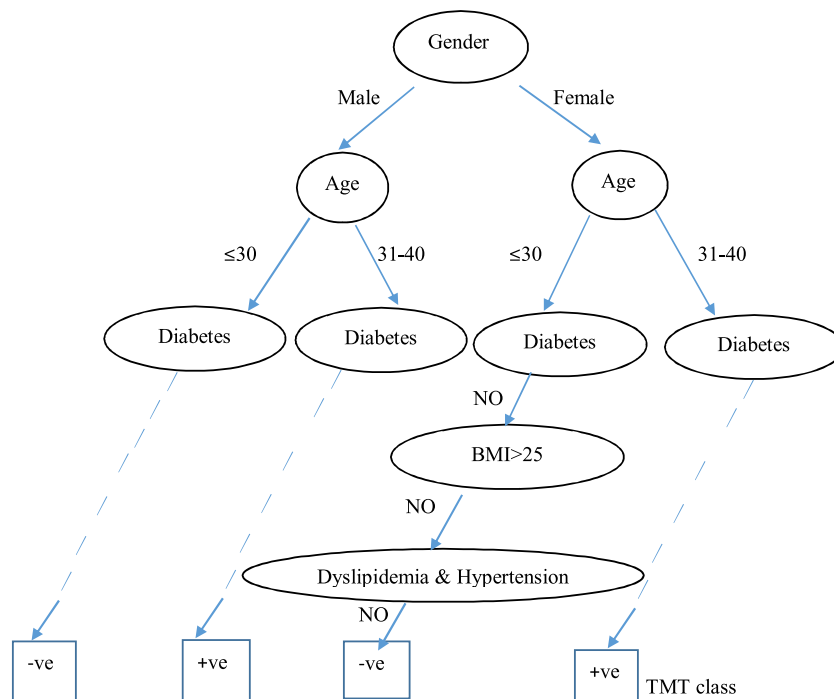

Fig. 2. Decision Tree.

2.1.4. Implementation of Decision Tree Method

The Decision Tree method[13] was applied on the Classified and Clustered dataset which was converted to Classification IF-THEN rules by tracing the path from the root node to each leaf node in the tree (Fig. 2). For example one of the rules extracted from the tree is:

R1: IF *Gender = Female* & *Age ≤ 30* & *Dyslipidemia = no* & *Diabetes = no* &*BMI < 25*THEN

TMT class = 'Negative'

From Table 1, it is evident that the application of rules of Decision Tree yields 73% Accuracy, which can be improved. With the scope of increasing the Accuracy, the analysis has been advanced to device an alternate algorithm and the search has been extended to find the procedures for the exact prediction of TMT result.

2.1.5. Pattern Recognition

In order to improve the Accuracy, the Pattern Recognition strategy has been proposed. Comparing the new patient's data with the existing information helps in increasing the accurate detection of the illness. With respect to Data Mining Pattern Recognition procedure, KNN method[14] was chosen and applied in our research for the prediction of TMT. KNN with its characteristic low Accuracy& time complexity[15] leads to a revised procedure in terms of 'Divide and Conquer' method. This algorithm is named as 'K-Sorting & Searching (KSS)'.

2.1.5.1. Implementation of K-Sorting & Searching (KSS) Algorithm:. While handling large amount of data, searching for a particular pattern in large database is a time consuming process.

| E Age	F BMI	G TMT	H Calculated TMT	I Age	J Hypertension	K TMT	L Calculated TMT	M Age	N Dyslipedemia	O TMT	P Calculated TMT	Q Age	R Diabetes	S TMT	T Calcualted TMT	U Sum of Calcuated TMT
21	22.4	0	0.067648	21	0	0	0.10427	21	0	0	0.29749	21	0	0	0.06719	0.5365976
23	24	0	0.001974	23	0	0	0.03941	23	1	0	0.20973	23	0	0	0.01677	0.267884
25	32.4	0	0.016012	25	0	0	0.02545	25	1	0	0.29635	25	0	0	0.03365	0.3714624
25	32.6	0	0.014378	25	0	0	0.02545	25	0	0	0.12425	25	0	0	0.03365	0.1977276
27	36	0	0.069286	27	1	0	0.26851	27	0	0	0.03763	27	0	0	0.08407	0.459496
29	25.65	0	0.236587	29	0	0	0.15517	29	1	0	0.46959	29	0	0	0.13449	0.9958369
30	30.94	1	0.234696	30	0	1	0.1876	30	1	1	0.5129	30	0	1	0.1597	1.09489644
30	26.64	0	0.269845	30	0	0	0.1876	30	0	0	0.0923	30	0	0	0.1597	0.70944464
30	27.33	0	0.264205	30	1	0	0.3658	30	0	0	0.0923	30	0	0	0.1597	0.88200458
30	27.1	0	0.266085	30	0	0	0.1876	30	0	0	0.0923	30	0	0	0.1597	0.7056846
30	39.7	0	0.163092	30	0	0	0.1876	30	0	0	0.0923	30	0	0	0.1597	0.6026922
30	26.9	1	0.267719	30	1	1	0.3658	30	1	1	0.5129	30	1	1	0.5	1.6464194
30	27.3	0	0.26445	30	0	0	0.1876	30	0	0	0.0923	30	1	0	0.5	1.0443458
32	26.87	0		32	0	0		32	0	0		32	0	0		
32	31.3	0		32	0	0		32	1	0		32	1	0		
32	21.8	0		32	0	0		32	0	0		32	0	0		
33	30.84	0		33	0	0		33	1	0		33	0	0		
33	25.8	0		33	0	0		33	1	0		33	0	0		
33	30.8	0		33	0	0		33	0	0		33	0	0		
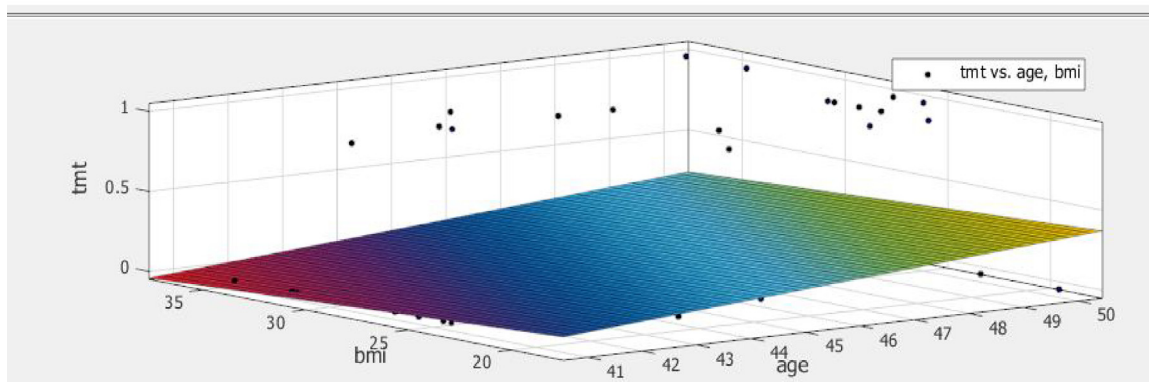
Fig. 3. An illustration for the mathematical calculation of TMT value &The Curve Fitting for TMT with Age and BMI.

For the efficient recognition of the pattern, KSS algorithm can be applied which is based on "Divide and Conquer" method. The Algorithm is as follows

K-Sorting & Searching (KSS) algorithm

1. Label the most significant numeric attribute as SORT_FACT.
2. Arrange the records in the TRANING data set in ascending order of SORT_FACT.
3. Divide the TRANING data into different ranges based on SORT_FACT.
4. Assign k value for each range in such a way that k>=0.
5. Store the first record SEQ_NO in FIRST_POINT[k] and the last record SEQ_NO in LAST_POINT[k] for each category of k value.
6. Get the TEST data
7. Assign k value for the TEST data based on the range defined by SORT_FACT.
8. Call DIS_CAL (TRAINING, TEST) for TRAINING = FIRST_POINT[k] to LAST_POINT[k]
9. TRAINING record which has minimum distance with the TEST record will be considered as the most similar search result and the corresponding pattern will be the required output.
10. End.

DIS_CAL (TRINING, TEST)

1. Find the Hamming distance (D_H) between the attributes of TRAINING and TEST data.[16]
a) Let x_i be an attribute of a TEST data and y_i be an attribute of the TRAINING data where i = 1,2,3 . . . then the Hamming distance between x_i and y_i is as given in Eq. (2).

$$D_H = \sum_{i=1}^{k}|xi - yi| \qquad (2)$$

where $x_i = y_i => D_H (x_i,y_i) = 0$

$x_i \neq y_i => D_H (x_i,y_i) = 1$

While applying KSS algorithm on the TMT dataset, the variable SORT_FACT was identified as age, which is a most significant attribute in the given dataset and all the patients' details were sorted based on SORT_FACT (age). The ranges were defined based on SORT_FACT and the k value was assigned for each range as follows, If age ≤ 30 then k = 0; If age = 31–40 then k = 1; If age = 41–50 then k = 2; If age = 51–60 then k = 3; If age ≥ 60 then k = 4.

Let us consider the TEST data with Age = 45. Unlike KNN, in this method it is not necessary to search the pattern in the entire training data. Instead, according to the KSS algorithm, k was assigned a value of 2 as in step 7 of KSS algorithm. The Hamming distance (Eq. (2)) was calculated between the TEST data and the TRAINING data which is in the range of FIRST_POINT[2] to LAST_POINT 2 as described in step 8 of KSS algorithm. The TRAINING data which has minimum distance with the TEST data in the range k = 2 was taken as the search result. The TMT value corresponding to the TRAINING data was the predicted TMT result of the TEST data. MATLAB software was used for programming purpose to test this procedure before its usage in the app. The results obtained reveal the time complexity of O (n log n).

As listed in Table 1, the Accuracy was 78% when KSS was applied on TMT Prediction which is an improvement over Decision Tree.

2.1.6. Implementation of Curve Fitting Method

The mathematical derivations such as polynomial methods were used to identify the relationship with simple empirical models.[17] This was applied on the existing system in order to find the mathematical relationship between the attributes for the prediction of TMT result.

Polynomial models for Curves are given by Eq. (3) as cited in.[18]

$$y = \sum_{i=1}^{n+1}p_i x^{n+1-i} \qquad (3)$$

where n + 1 is the order of the polynomial, n is the degree of the polynomial where $1 \leq n \leq 9$.
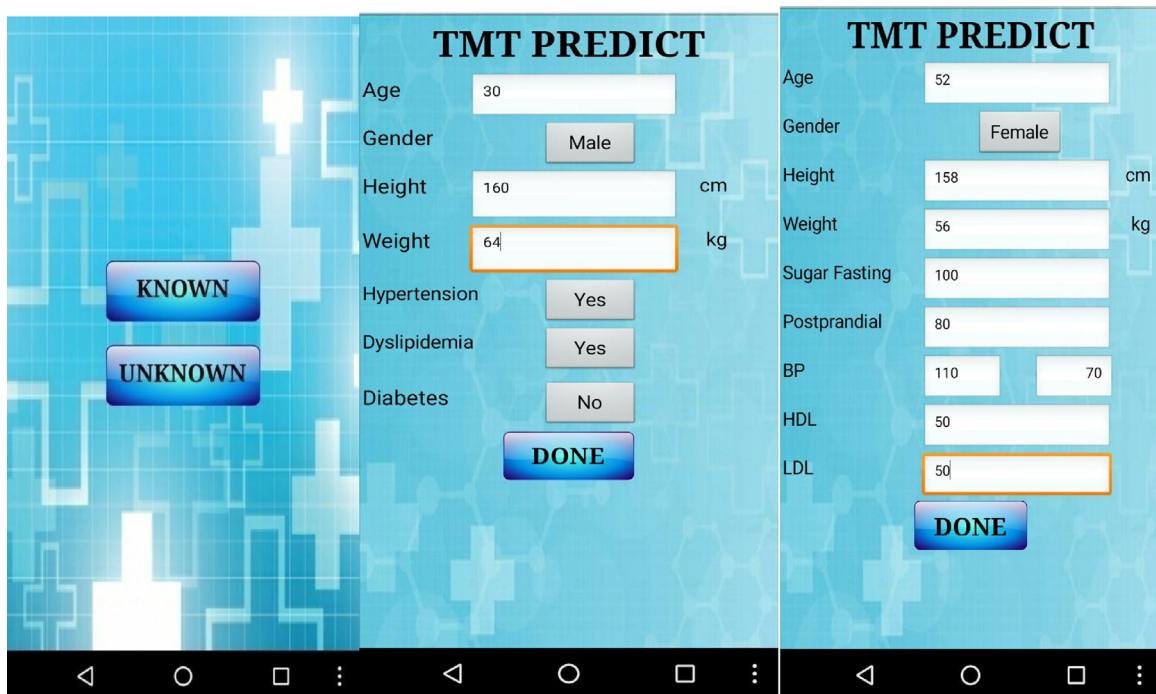

Fig. 4. Screen Shots of the Mobile app "TMT Predict" for Known and Unknown Cases.

In order to derive the Polynomial function, the attribute combination of age and TMT, with respect to BMI, systemic hypertension, dyslipidemia and diabetes, were considered individually through Curve Fitting methodology, and executed with the help of MATLAB Curve Fitting tool. The Polynomial derived for the attribute combination of age and TMT with respect to BMI is given in Eq. (4).The Goodness of Fit for the polynomial was calculated in terms of Sum of Squared Errors (SSE), R-Square, Adjusted R-Square and Root Mean Square Deviation (RMSE).The corresponding values are as in the list given below. Similarly other polynomial functions were retrieved for other attribute combinations.

$$f(x, y) = p00 + p10 * x + p01 * y \qquad (4)$$

where f(x,y) represents TMT
   x represents Age
   y represents BMI
   p00, p10 & p01 represent coefficients

Coefficients (with 95% confidence bounds):
p00 = − 0.919 (−2.126, 0.2878)
p10 = 0.02791 (0.003078, 0.05274)
p01 = − 0.006745 (−0.02655, 0.01306)
Goodness of fit:
Sum of Squared Errors (SSE): 14.27
R-square: 0.04722
Adjusted R-square: 0.0289
Root Mean Square Deviation (RMSE): 0.3704

The derived polynomial functions were reduced to a single polynomial equation with all the attributes for the exact prediction of TMT. The resultant TMT values are analyzed as shown in Fig. 3. This reveals that for instance, in the case of patients whose age group is less than 30 and if the calculated TMT value is less than 1, then the TMT result of the model will be taken as Negative. Otherwise it would be Positive (for age group ≤30). Similarly, the TMT range has been fixed for other age groups to classify the Positive and Negative TMT. The result analysis, in Table 1, shows

Table 1
Results of 250 test data for Decision Tree, Curve Fitting, KSS Pattern matching and the proposed TMT Predict along with its Accuracy Graph.

|		Age Group						
		≤30	>30 & ≤40	>40 & ≤50	>50 & ≤60	>60 & ≤70	≥ 70	Overall
Decision Tree	TP	0	0	6	24	6	8	**44**
	TN	11	33	56	27	9	2	**138**
	FP	0	3	17	8	6	1	**35**
	FN	1	1	7	14	9	1	**33**
	Se %	0	0	46	63	40	88	**57**
	Sp %	100	92	77	77	75	66	**80**
	Acc %	92	89	72	70	50	83	**73**
Curve Fitting	TP	1	1	8	31	11	7	**59**
	TN	10	27	66	29	12	3	**147**
	FP	1	9	7	6	3	0	**26**
	FN	0	0	5	7	4	2	**18**
	Se %	100	100	62	82	73	78	**77**
	Sp %	90	75	90	83	80	100	**85**
	Acc %	92	82	86	82	77	83	**82**
KSS Pattern matching	TP	0	0	6	25	8	5	**44**
	TN	11	33	68	26	11	3	**152**
	FP	0	3	5	9	4	0	**21**
	FN	1	1	7	13	7	4	**33**
	Se %	0	0	46	66	53	55	**57**
	Sp %	100	92	93	74	73	100	**88**
	Acc %	92	89	86	70	63	66	**78**
TMT Predict	TP	0	0	6	28	10	7	51
	TN	11	35	70	28	11	3	**158**
	FP	0	1	3	7	4	0	**15**
	FN	1	1	7	10	5	2	**26**
	Se %	0	0	46	74	67	78	**69**
	Sp %	100	97	96	80	73	100	**91**
	Acc %	92	95	88	77	70	83	**84**
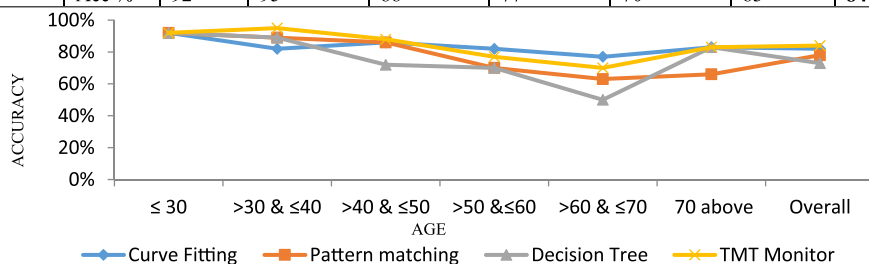
Table 2
The real time result for 300 patients through "TMT Predict" app.

Real time Analysis	Evaluative Metrics	I	II	III	IV	V	VI	Total
Mobile App- "TMT Predict"	TP	2	5	6	2	4	5	**24**
	TN	33	29	31	25	29	32	**179**
	FP	10	9	6	10	10	9	**54**
	FN	5	7	7	13	7	4	**43**
	Se%	29	42	60	15	36	55	**35**
	Sp%	77	76	84	71	74	78	**77**
	Acc%	70	68	74	54	66	74	**67**
Angio	TP	2	4	7	2	3	6	**24**
	TN	33	31	33	26	32	33	**188**
	FP	5	5	5	12	4	3	**34**
	FN	10	10	5	10	11	8	**54**
	Se%	20	29	58	17	21	43	**35**
	Sp%	87	86	94	68	89	92	**80**
	Acc%	70	70	80	56	70	78	**71**

that the Accuracy value of this procedure is found to be 82%, which is 9% increase from the Decision Tree methodology Fig. 4.

2.1.7. TMT Predict

The three methods were applied to develop a simple user-friendly Android App called " TMT Predict". The app "TMT Predict", predicts the results of TMT for each of the three methods independently and compares the results. If the results of the Decision Tree, Curve Fitting and Pattern recognition methods are identical then "TMT Predict" will also deliver the same result. Otherwise, the two leading results out of the three methods can be treated as the final TMT result of the app. For example, if the result of TMT in case of KSS and Curve Fitting are positive but, Decision Tree is negative then the result of "TMT Predict" will be taken as positive. Compared to the three methods implemented individually, the results of the "TMT Predict", obtained using the combination of three methods, were higher and has been implemented in the Android mobile application. The Accuracy has been tested in real time environment and will be uploaded in the Google Play Store for the patients' usage.

3. Evaluation methods

Evaluation of a model is the integral part of developing a Classifier/Predictor.[19,20] The app was tested in real time with 300 new patients, and the positive, negative predictive values and the accuracy in predicting the result of the TMT were calculated from True Negative (TN), True Positive (TP), False Negative (FN) and False Positive(FP) values.

Accuracy = (TN + TP)/(TN + TP + FN + FP).
Positive Predictive Value = TP/(TP + FP).
Negative Predictive Value = TN/(TN + FN).

4. Results & discussions

The three different methods Decision Tree, KSS and Curve Fitting, along with the "TMT Predict" app, were assessed in the test set of 250 patients. The results are shown in Table 1. The Accuracy of the Decision Tree, KSS and Curve Fitting was 73%, 78% and 82% respectively. The "TMT Predict" app which combined all the three methods improved the accuracy to 84%.

The results were analyzed for different age categories as shown in Table 1. The accuracy of the app when the age was greater than 60 years decreased from 84% to 76%. This raises the question of other influencing factors like retirement stress which may have to be considered in this age group. This opens up a new avenue for future research. When the working age category (age between 30 years and 60 years) was considered, the accuracy of the "TMT Predict" app improved from 84% to 88%. This behavior of the dataset paves the way for further inclusion of other compounding risk factors in different age groups, which is the future scope of this work. The flexibility of the app is that it can be modified with new attributes in subsequent versions.

As the final step, the app was tested prospectively on 300 new patients. The patient count was taken as 50 for each group and the results are tabulated in Table 2. The sensitivity and specificity of the "TMT Predict" app were 35% and 77%, respectively. Based on the TMT Predict, the Coronary Angiographic results for the same set of patients were also analyzed, which showed a specificity of 80%, whereas the sensitivity remained the same at 35%. This indicates that the TMT Predict app is more specific than the TMT itself in predicting the results of coronary angiography. For the real time analysis, the Positive and Negative Predictive values were calculated.[21] The Positive Predictive Value and Negative Predictive Values for the TMT were 40% and 83% and for coronary angiogram, it was 52% and 80%. The Negative Predictive value of the app was high, indicating that it is a good screening tool to rule out CAHD.

There are no similar datasets in literature using multiple Data mining techniques to predict the results of TMT.

5. Conclusion

The "TMT Predict" app is a novel method to predict the result of TMT. It is a simple, user-friendly app requiring only six parameters (age, sex, BMI, diabetes, dyslipidemia, and systemic hypertension). Utilizing Data Mining techniques (Decision Tree, KSS method based on KNN algorithm), and the Curve Fitting method, the app was able to give a very high Negative Predictive value although the sensitivity was not satisfactory. This makes the app an useful tool to "rule out" CAHD. The data analysis has also given interesting outputs like reduced Accuracy if the age is more than 60 years, indicating that there are other factors that will have to be analyzed in the future. The methodologies used in this app are simple and flexible giving scope for future refinement. The proposed "TMT Predict" app can be a digital replacement for the manual TMT in the future.

References

1. http://www.who.int/mediacentre/factsheets/fs310/en/index2.htmlblood.
2. http://www.thehealthsite.com/diseases-conditions/heart-disease-in-india-6-shocking-facts-you-should-know/.
3. http://food.ndtv.com/health/why-is-india-experiencing-a-cardiovascular-disease-epidemic-771837.
4. Saraee Mohamad, Waheed Asad, Javed Sargheel, Nigam Joshi. Application of data mining in medical domain: case of cardiology sickness level. :337.

5. Lavekar Anurag S, Salkar Harsha R. Treadmill test to detect stress induced ischemic heart disease in type 2 diabetes mellitus patients asymptomatic for CAD: a hospital based cross-sectional study in rural population of central India. *J Diabetes Metab*. 2013;4(2):1000244.

6. Alizadehsani Roohallah, Habibi Jafar. Mohammad javad hosseini, hoda mashayekhi, reihane boghrati, asma ghandeharioun, behdad bahadorian, and zahra alizadeh sani a data mining approach for diagnosis of coronary artery disease. *Comput Methods Programs Biomed*. 2013;111(1):52–61.

7. Chaurasia Vikas, Pal Saurabh. Data mining approach to detect heart diseases. *Int J Adv Comput Sci Inf Technol (IJACSIT)*. 2014;2:56–66.

8. Tsipouras Markos G, Exarchos Themis P, Fotiadis Dimitrios I, et al. Michalis Automated diagnosis of coronary artery disease based on data mining and fuzzy modeling. *IEEE Trans Inf Technol Biomed*. 2008;12(4):447–458.

9. Podgorelec Vili, Kokol Peter, Stiglic Milojka Molan, Heričko Marjan, Rozman Ivan. Knowledge discovery with classification rules in a cardiovascular dataset. *Comput Methods Programs Biomed*. 2005;80:S39–S49.

10. Yang Yanga, Faridb Suzanne S, Thornhilla Nina F. Data mining for rapid prediction of facility fit and debottlenecking of biomanufacturing facilities. *J Biotechnol*. 2014;179(2014):17.

11. Fan Cheng, Xiao Fu, Wang Shengwei. Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. *Appl Energy*. 2014;127:1–10.

12. Jerline Amutha A, Padmajavalli R. Classification of cardiology data for the prediction of treadmill test. *Int J Control Theory Appl*. 2017;10(36):231–243.

13. Kadi Ilham, Idri Ali, Fernandez-Aleman JL. Knowledge discovery in cardiology: a systematic literature review. *Int J Med Inf*. 2017;97:12–32.

14. Shouman Mai, Turner Tim, Stocker Rob. Applying k-nearest neighbour in diagnosing heart disease patients. *Int J Inf Edu Technol*. 2012;2(3):220.

15. Jiang Liangxiao, Cai Zhihua, Wang Dianhong, Jiang Siwei. Survey of improving k-nearest-neighbor for classification. *Fuzzy Systems and Knowledge Discovery, 2007. FSKD 2007. Fourth International Conference on*. 2007;vol. 1:.

16. Pandit Shraddha, Gupta Suchita. A comparative study on distance measuring approaches for clustering. *Int J Res Comput Sci*. 2011;2(1):29.

17. https://in.mathworks.com/help/matlab/ref/polyfit.html.

18. https://in.mathworks.com/help/curvefit/polynomial.html.

19. Raikwal JS, Saxena Kanak. Performance evaluation of SVM and k-nearest neighbor algorithm over medical data set. *Int J Comput Appl*. 2012;50(14).

20. Baratloo Alireza, Mostafa Hosseini, Ahmed Negida, El Gehad Ashal. Part 1: simple definition and calculation of accuracy, sensitivity and specificity. *EMERGENCY-An Acad Emergency Med J*. 2015;3(2):48–49.

21. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol*. 2008;56(1):45–50.